

Citation for published version:

Faraway, J 2016, 'Confidence bands for smoothness in nonparametric regression', *Stat*, vol. 5, no. 1, pp. 4-10.
<https://doi.org/10.1002/sta4.100>

DOI:

[10.1002/sta4.100](https://doi.org/10.1002/sta4.100)

Publication date:

2016

Document Version

Early version, also known as pre-print

[Link to publication](#)

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Confidence Bands for Smoothness in Nonparametric Regression

Julian Faraway

13 November 2015

Abstract

The choice of the smoothing parameter in nonparametric regression is critical to the form of the estimated curve and any inference that follows. Many methods are available that will generate a single choice for this parameter. Here we argue that the considerable uncertainty in this choice should be explicitly represented.

The construction of standard simultaneous confidence bands in nonparametric regression often requires difficult mathematical arguments. We question their practical utility, presenting several deficiencies. We propose a new kind of confidence band that reflects the uncertainty regarding the smoothness of the estimate.

keywords: Gaussian process regression; smoothing parameter; Bayes estimation; confidence interval.

1 Introduction

Simultaneous confidence bands (SCB) are a popular topic in the statistical research literature. There is a long history of research in the area dating back to the work of Scheffé (1959) on confidence bands for parametric regression. The greatest volume of research concerns the nonparametric regression problem where we are given data pairs, $(x_1, Y_1), \dots, (x_n, Y_n)$ which are generated from the model

$$Y_i = f(x_i) + \epsilon_i$$

where f is an unknown function. We are asked to construct a pair of bands, $(UCB(x), LCB(x))$ such that

$$P(LCB(x) \leq f(x) \leq UCB(x) \quad \forall x \in I) = p$$

where p is typically 0.95. There are numerous variations of the basic problem concerning the assumed smoothness of f , the type of estimator used, the nature of the error ϵ and the domain I of the band. The problem has been extended to higher dimensions and used for binary response regression, survival analysis and other regression-like methods. Nonparametric regression is used as a tool within more general methods like generalized additive models and SCBs have been constructed within.

Highly cited papers in the area include Härdle and Bowman (1988), Hall and Titterton (1988) and Eubank and Speckman (1993). Searches of scholarly databases reveal hundreds of papers on this topic. The problem has attracted extensive and sustained interest because of the intriguing mathematical challenges it poses. No solution has been found completely satisfactory and so new publications appear regularly.

Examples of confidence bands are shown in Figure 1. The data come from observations of the brightness (magnitude) of a distant galaxy which has an active galactic nucleus. The

data were collected by the Catalina Real-time Transient Survey (Drake et al. (2009)). Higher magnitudes correspond to lower brightness so the vertical scale is inverted according to standard astronomical practice. The timings of the observations are not uniform because sometimes the galaxy cannot be observed due to the seasonal movement of the Earth so seven larger clusters can be seen. The telescope records up to four observations within a 30 minute period so there are finer clusters which are not immediately obvious in the plot.

The first plot shows the default bands supplied by the popular `ggplot2` R package of Wickham (2009). The smoother used is Lowess. Although it is not immediately apparent, this fit is substantially oversmoothed. In the second plot, a cubic spline-based smoother from the `mgcv` package of Wood (2006) is displayed. Here the amount of smoothness has been selected using GCV. We use this as a continuing example below.

This article discusses the drawbacks of SCBs and questions their utility in statistical practice in §2. If we accept the deficiencies of the standard bands, we might seek something more informative which we present in §3. We conclude with a discussion in §4 with some comparison to other alternatives.

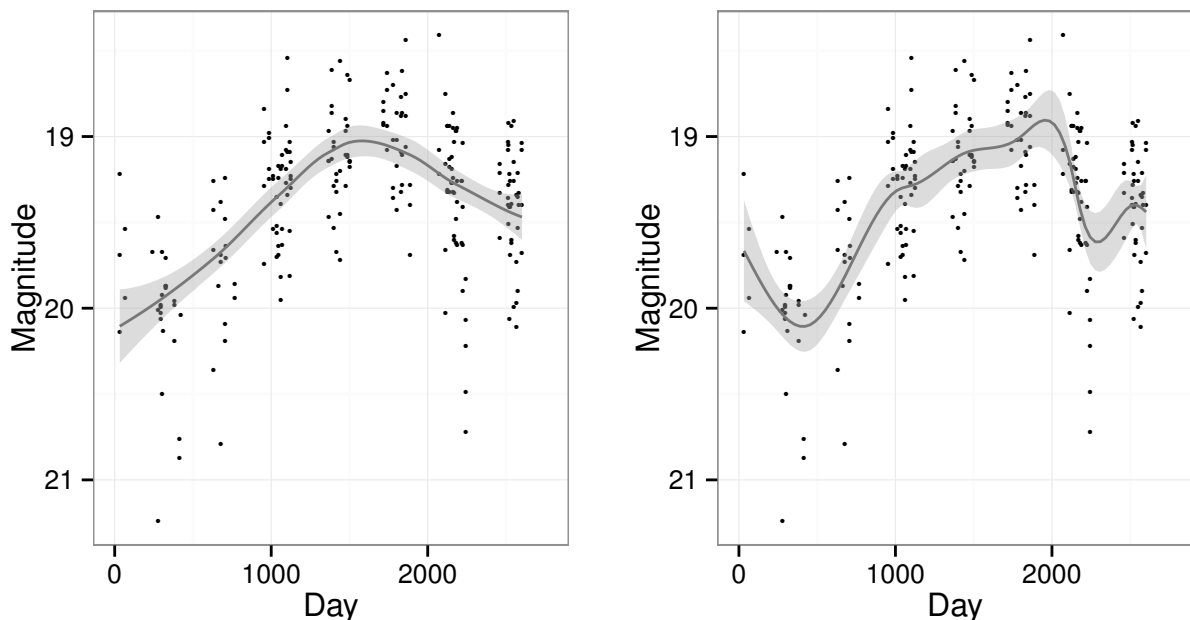


Figure 1: 95% Confidence bands for a light curve for an active galactic nucleus. Default Lowess based bands are shown on the left and spline-based bands with automated smoothing selection on the right.

2 Deficiencies of the standard bands

What is the practical purpose of SCBs? Their definition is clear but this is not, in of itself, proof of their utility. Let us suppose we are satisfied with the technical mechanics of their construction. What conclusion will we make or what decision regarding the subsequent analysis will we take on seeing the SCB?

Many users might want an informal graphical expression of uncertainty in the curve estimate. For this purpose, they are not concerned with the precise numerical accuracy of the bands shown on the plot. Consider the examples seen in Figure 1. The bands supplied are pointwise rather simultaneous. This is not much of a problem since, barring exceptional circumstances, SCBs will be the same shape and just a little wider than the pointwise bands. Indeed, many bands take the form $\hat{f}(x) \pm c.se(\hat{f}(x))$. The value of c is a little larger for the simultaneous rather

than pointwise version of the bands. For the user who seeks only an approximate graphical representation of the uncertainty, the distinction is not important since no formal conclusions will be made. Now SCBs would be a somewhat better choice in these circumstances, but pointwise bands are much easier to calculate. Many smoothing methods are locally linear so calculating the standard error at a point is typically not difficult and hence constructing the pointwise bands is simple enough. In contrast, SCBs often require complex calculations or time-consuming methods like bootstrapping. This is more effort than is justified for informal use.

In order to justify the additional effort in constructing accurate SCBs, we must have some specific interest in their numerical values. We might imagine a future user who is interested in simultaneously testing a set of hypotheses of the form: $H_0 : f(t_i) = f_i$ where the particular values t_i are unknown to us in advance. Certainly, the SCBs would allow us to perform these tests accurately with the correct overall size but it seems very unlikely that anyone would want to do this. Certainly, there are applications involving such multiple comparisons but these typically arise in factorial experiments and not regression problems.

The formal user of SCBs will almost certainly seek information about all of f and not specific points along f . One common question will be “What is a plausible parametric functional form for f ?” The most likely form for such a question would be “Is f a straight line?” It is easier to work with parametric functions for prediction and explanation purposes so there is ample motivation for the analyst to change the model in this way. Even so, the analyst wishing to convert or simply test a parametric form will find it easier and statistically more powerful to embed the proposed form within a more general parametric family and test accordingly. Although it would be valid to simply check whether the proposed parametric form fits within the SCBs, this is not the best approach because SCBs may get the size right but the power will be lacking as we discuss later.

The other common question that might make use of SCBs is “Does f have a particular qualitative shape?” In statistics research articles on SCBs where a data example is presented, the interest is often in whether some secondary peak seen in \hat{f} is real. SCBs can be used to answer this question but are not the best tool. More powerful specific tools have been developed that can be used to detect such peaks, for example the SiZer method of Chaudhuri and Marron (1999). In particular, the presence of peaks in \hat{f} is highly sensitive to the choice of the smoothing parameter. Unfortunately, SCBs typically only reflect the uncertainty in the amplitude of f and not its smoothness.

We claim that SCBs are an inefficient answer to the questions that analysts are likely to ask. There are much better specific solutions available. There is a general difficulty with SCBs which make them less useful for formal testing. When confidence intervals are provided, there is a sense in which any parameter value that falls within the interval is plausible. In many cases, the value corresponds to a null hypothesis that would not be rejected. Yet this notion of plausible values within the intervals does not extend to SCBs.

Consider the two examples shown in Figure 2. In the first example, the proposed curve fluctuates greatly but remains within the bands. Depending on the application, we may have some opinions about how rough or smooth f is likely to be but the SCBs add nothing to the debate. The proposed curve fits within the bounds so it represents an acceptable null hypothesis but we do not find this helpful in determining its plausibility. Curves ranging from very smooth to very rough fit between the bands.

In the second example, the proposed curve takes the form $UCB(x) - \delta$ for a small value of δ . Although the curve falls entirely within the bands, it is implausible. Consider a null hypothesis of the form $H_0 : f = \hat{f}$ against alternatives of the form $H_1 : f = \hat{f} + \gamma$ for values of $\gamma \neq 0$. Since we have fixed the shape as \hat{f} , a test merely needs to estimate γ . The situation reduces to a simple t-test on the data $Y_i - \hat{f}(x_i)$. We can see that the proposed curve will be rejected

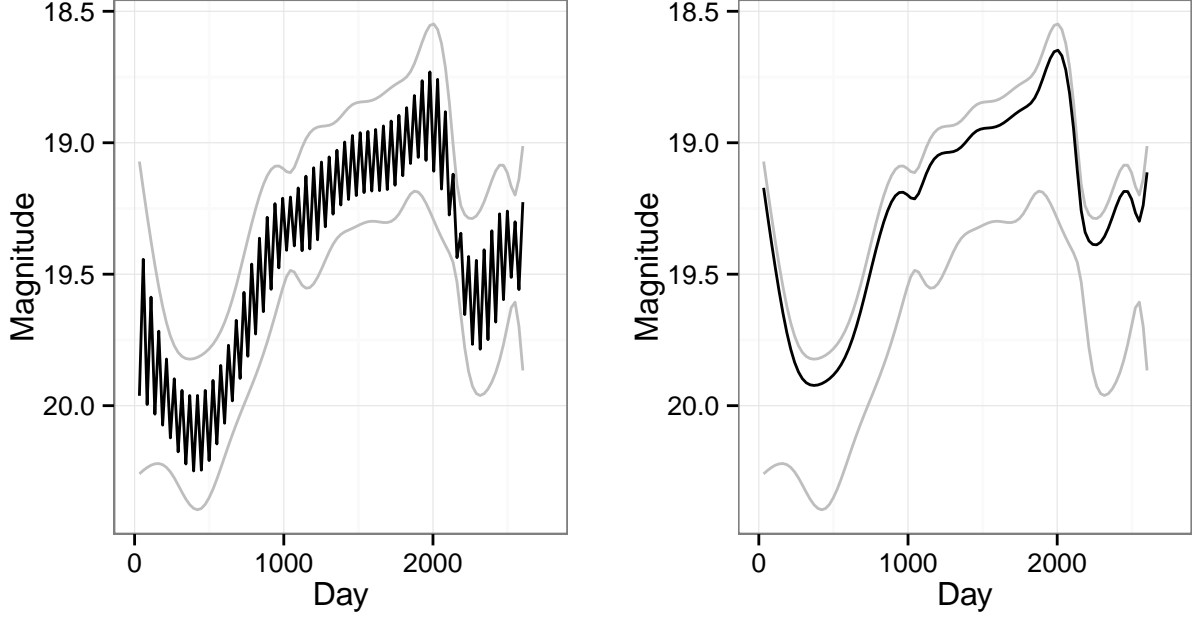


Figure 2: Implausible proposed curves that fit within the confidence bands.

since the estimated value of γ will be relatively large. This is because γ will be estimated using all the data whereas the $se(\hat{f}(x))$ will effectively only use observations local to x .

There will be a multitude of highly implausible proposed curves that fit within the bands while other more reasonable suggestions may escape the bands. The essential difficulty is that we must project the infinite dimensional function space into a low dimension. Information is inevitably lost in this process so tests based on SCBs will have problems with power.

The main uncertainty in this problem is amount of smoothing. As can be seen in Figure 1, the choice of smoothing parameter makes a big difference to the result. The width of the bands is a secondary matter. We need to focus on this. In the next section, we propose a different set of bands that relate to this primary uncertainty.

3 Confidence bands for smoothness

We use Gaussian process regression (GPR) to construct a graphical depiction of the uncertainty in the smoothness of the bands. The method is Bayesian but could be adapted to other types of smoother such as splines. We have a Gaussian process: $f(x) \sim GP(\psi(x), k(x, x'))$ with mean $\psi(x)$. We use the popular squared covariance kernel:

$$k(x, x') = \sigma_f^2 \exp\left(-\frac{1}{2l^2}(x - x')^2\right) + \sigma_n^2 \delta(x - x') \quad (1)$$

where $\delta(z)$ is 0 when $z \neq 0$ and 1 when $z = 0$. Other choices of kernel are possible and could easily be substituted here. See Rasmussen and Williams (2006) for a general introduction.

The three parameters, l , σ_f^2 and σ_n^2 , are strictly positive. We desire a prior for these parameters that gives most weight to small to moderate values but allows the possibility of larger values. In keeping with the recommendations of Polson and Scott (2012), we use a half-Cauchy prior of the form $C^+(0, \tau)$ where we must choose the scaling τ . In all, there are four components of the prior which must be specified. We have given concrete suggestions for the scaling of the priors but these could also be chosen using the usual subjective approach if desired.

1. σ_f^2 is the signal variance. We set σ_f^2 to follow half-Cauchy prior $C^+(0, \tau)$ with a scale $\tau = SD(Y)$.
2. σ_n^2 is the noise variance. This represents the measurement error. In our particular example, we have a good estimate of what this is likely to be due to Astronomers knowledge of their instrumental accuracy. For generality, we abstain from using this helpful information and use a half-Cauchy prior with scaling τ set to the standard deviation of the absolute differences between successive Y .
3. l is sometimes called the length-scale. It controls the amount of correlation between parts of the curve and therefore the amount of smoothness in the resulting posterior fit. This is the most important parameter. We use a half-Cauchy prior with a scaling τ set to one quarter of the range of x .
4. $\psi(x)$ is the prior mean. We set this to be constant at the median response value. In other examples, we might use some simple parametric form such as a line. In part this choice is governed by how we would like the curve to extrapolate beyond the range of the data. If we don't care much about this, then any simple choice would be appropriate.

The half-Cauchy prior distribution used for three of the parameters sets the broad expectation for the result but the thick tail of this distribution permits the possibility of an outcome significantly different from prior expectations. The outcome is insensitive to the particular choices of the τ s and we merely need to get a sensible order of magnitude for these values as the prior is not sharp.

The Bayes computation was implemented in Stan Development Team (2015). The code for Gaussian process regression is a short example in the STAN reference manual so it was simple to implement. MCMC computation took about two minutes on the author's aging computer for the data here. For larger datasets, some economy may be achieved using methods discussed in Quiñonero Candela and Rasmussen (2005) and elsewhere. Our implementation is perhaps too slow for rapid exploratory work where the analyst may want to check many plots quickly but certainly fast enough for relationships we care about specifically.

In principle, we can answer the questions of interest by consulting the posterior distribution but we need more accessible representations. Our main interest is in the length-scale l that controls the smoothness of the curve. We can use the samples from the posterior to calculate a credible interval (l_L, l_U) of a chosen level for this parameter but the numbers are difficult to interpret. We propose the following method to gain intuition:

1. Find the posterior medians for σ_f^2 and σ_n^2 .
2. Compute the posterior mean of $f(x)$ given these values and setting $l = l_L$. This will be the rough band.
3. Repeat but setting $l = l_U$. This will be the smooth band.

The computation of the posterior given the three parameters and the prior mean is explicit and rapid as described in Rasmussen and Williams (2006). The resulting bands for the data are shown in Figure 3.

The plot gives a visual impression of how rough or smooth the relationship is. This is unlike the usual SCB where the statement is about the function lying between the bands. In this particular example, we see the smooth band still has eleven optima so we are unlikely to find a satisfactory simple parametric form for this relationship. The rough band has 35 optima so we are quite confident that the relationship has a considerable number of peaks and valleys. In the astronomy application, the purpose is to classify new light curves into types such as supernovae

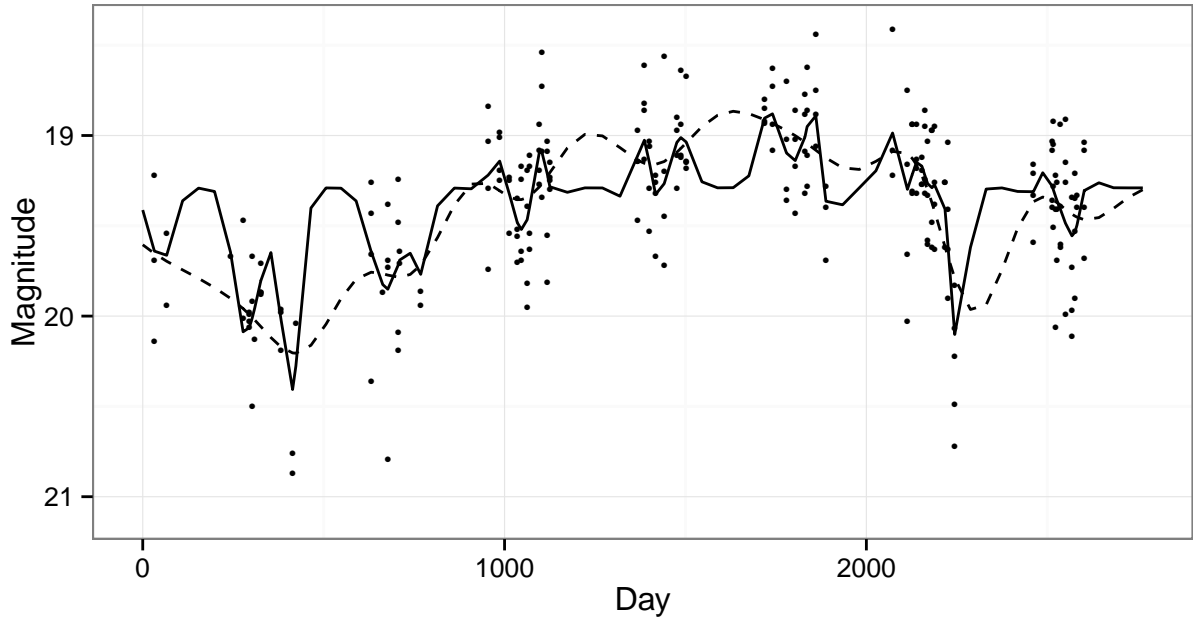


Figure 3: Confidence bands for smoothness. The two curves represent the posterior means for the upper and lower ends of a 95% credibility interval for the smoothing parameter with other parameters set at their posterior median values.

or blazars etc. The shape of the curve is crucial in making the classification so the bands here are useful. The usual SCB would not be very helpful — in particular, our conclusions would be very different judging only from Figure 1.

4 Discussion

The SiZer method of Chaudhuri and Marron (1999) was developed to answer questions about the existence of local optima. It has since been extended to answer several other questions like the existence of jumps. The method recognises that the amount of smoothing can be critical is deciding whether a particular feature exists and produces a diagram that describes how some measure of significance at each point in the range of interest varies as the smoothing parameter varies. This is very helpful but SiZer avoids the extreme of picking a single value for the smoothing parameter to being agnostic about its value. The method proposed in this article is different in that I propose that attention be focused on the most likely values of the smoothing parameter while remaining agnostic about what features are of interest.

The smoothness bands for one of the examples discussed in Chaudhuri and Marron (1999) are presented in Figure 4. The data come from Bralower et al. (1997) who reported the ratio of strontium isotopes found in fossil shells in the mid-Cretaceous period from about 90 to 125 million years ago. The strontium ratio has been rescaled in the same manner as the SiZer paper. We can see that the shape of the curve is not much in doubt for the older half of the curve but there is substantially more uncertainty for the more recent period. In the SiZer paper, the apparent dip around 97 million years was deemed insignificant but in this plot we can see the dip is barely present in the smooth band and clearly present in the rough band. Hence, we see some evidence for a dip. A more definitive answer would require an investigation of the full posterior which lies beyond the purpose of this article.

In Erästö and Holmström (2005), a Bayesian version of SiZer is developed using smoothing splines which does discuss placing a prior on the smoothing parameter. The main interest is

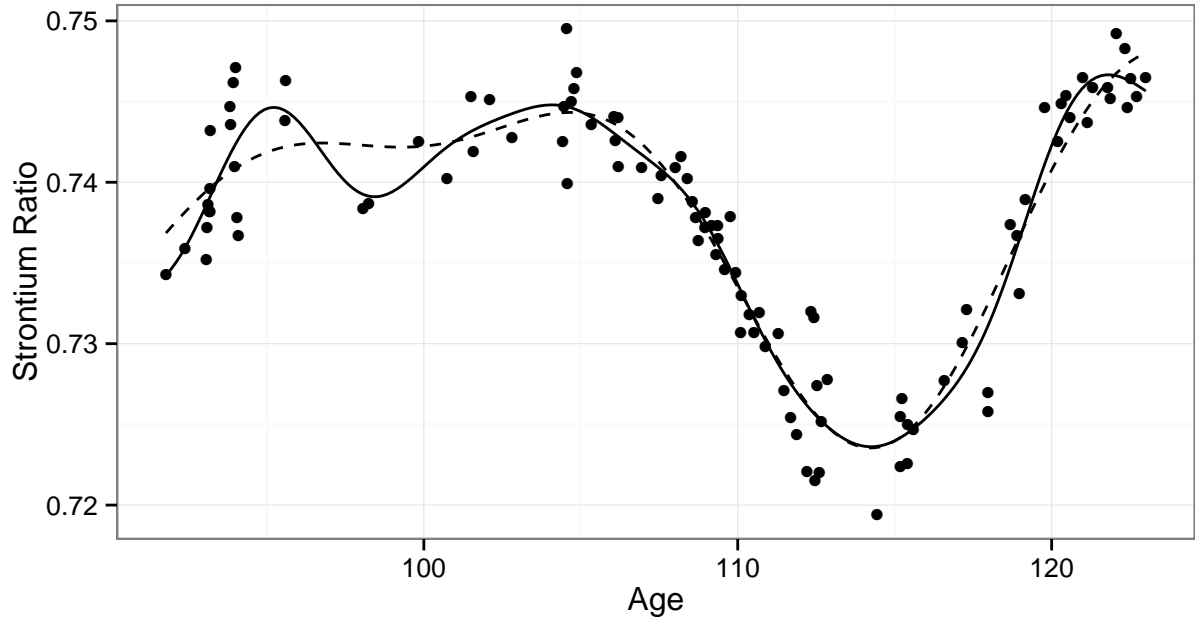


Figure 4: Smoothness bands for the fossil data. The solid curve is the posterior mean curve at the 2.5%ile of smoothness while the dashed band is at the 97.5%ile. The age is in millions of years and the strontium ratio has been linearly transformed from the original measurements.

in the development of the confidence bands which are used by SiZer rather than presenting the nature of the uncertainty regarding the smoothing parameter. Nevertheless, this does illustrate that a version of the method could be produced for a spline-based smoother rather than Gaussian process regression.

Further elaboration of the method presented here is possible. The idea could be extended to problems similar nonparametric regression such as density estimation. It would also be possible to combine the amplitudinal uncertainties of the original SCBs with the smoothness uncertainty but it would be difficult to display this in graphically accessible manner.

Appendix

Data and code to reproduce the examples shown here is available from people.bath.ac.uk/jjf23/scb

References

- Bralower, T., P. Fullagar, C. Paull, G. Dwyer, and R. Leckie (1997). Mid-cretaceous strontium-isotope stratigraphy of deep-sea sections. *Geological Society of America Bulletin* 109(11), 1421–1442.
- Chaudhuri, P. and J. S. Marron (1999). Sizer for exploration of structures in curves. *Journal of the American Statistical Association* 94(447), 807–823.
- Drake, A. et al. (2009). First Results from the Catalina Real-time Transient Survey. *The Astrophysical Journal* 696, 870–884.
- Erästö, P. and L. Holmström (2005). Bayesian Multiscale Smoothing for Making Inferences

- about Features in Scatterplots. *Journal of Computational and Graphical Statistics* 14(3), 569–589.
- Eubank, R. and P. Speckman (1993). Confidence bands in nonparametric regression. *Journal of the American Statistical Association* 88(424), 1287–1301.
- Hall, P. and D. Titterton (1988). On confidence bands in nonparametric density estimation and regression. *Journal of Multivariate Analysis* 27(1), 228–254.
- Härdle, W. and A. W. Bowman (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *Journal of the American Statistical Association* 83(401), 102–110.
- Polson, N. G. and J. G. Scott (2012, 12). On the half-cauchy prior for a global scale parameter. *Bayesian Analysis* 7(4), 887–902.
- Quiñonero Candela, J. and C. E. Rasmussen (2005, December). A unifying view of sparse approximate gaussian process regression. *J. Mach. Learn. Res.* 6, 1939–1959.
- Rasmussen, C. and C. Williams (2006). *Gaussian Processes for Machine Learning*. Cambridge: MIT Press.
- Scheffé, H. (1959). *The Analysis of Variance*. New York: Wiley.
- Stan Development Team (2015). *Stan Modeling Language Users Guide and Reference Manual, Version 2.8.0*.
- Wickham, H. (2009). *ggplot2: Elegant Graphics For Data Analysis*. New York: Springer.
- Wood, S. (2006). *Generalized Additive Models: An Introduction with R*. Boca Raton, FL: CRC press.